Burroughs-Wellcome Fund
Maryland Genetics, Epidemiology and Medicine (MD-GEM) Pre-doctoral Training Program

Abstract Book
Genetics Research Day

February 9, 2018

Contents

Dear Participants,

On behalf of the Maryland-Genetics, Epidemiology, Medicine Training Program (MD-GEM) it is our pleasure to welcome you to the fifth annual Genetics Research Day at Johns Hopkins University. MD-GEM includes faculty spanning the Mckusick-Nathans Institute of Genetic Medicine, the Johns Hopkins Bloomberg School of Public Health, the Johns Hopkins School of Medicine and the National Human Genome Research Institute, who join together to train doctoral students in population and laboratory sciences focused on genetics.

This Genetics Research Day provides the greater JHU community an opportunity to promote discussion and collaboration across JHU/NHGRI and to integrate students from different disciplines into the wide breadth of genetics research. We welcome all faculty, post-doctoral fellows and students, especially those new to the field of genetics. We look forward to continued partnerships and new relationships across the fields of Epidemiology, Biostatistics, Human Genetics, Biology, Computer Science, Mathematics and more.

The posters represent the Departments of Biostatistics, Epidemiology, and Mental Health of the Johns Hopkins Bloomberg School of Public Health; the Departments of Hematology, Neurology, Neuroscience, Oncology, Pathology, Psychiatry and Behavioral Sciences, and the Division of Allergy and Clinical Immunology in the Department of Medicine of the Johns Hopkins School of Medicine; the Berman Institute of Bioethics, Center for Computational Biology, Center for Epigenetics, Center for Inherited Disease Research, Lieber Institute for Brain Development, McKusick-Nathans Institute for Genetic Medicine, Sidney Kimmel Comprehensive Cancer Center, Welch Center for Prevention, Epidemiology & Clinical Research, and Wendy Klag Center for Autism of the Johns Hopkins University; the Division of Cancer Epidemiology and Genetics of the National Cancer Institute; and the Computational and Statistical Genomics Branch and Medical Genetics Branch of the National Human Genome Research Institute.

A very special thank you to Dr. Michael Boehnke, University of Michigan School of Public Health, for joining us as our plenary speaker. Thank you to all faculty judges who have generously lent us their expertise and time and to whom we are indebted. We extend our sincere thanks to Sandy Muscelli, Jon Eichberger and Nicole Thornton for all of their help in organizing and promoting this event. We are especially grateful for the tireless efforts of Jennifer Deal who graciously attended to every detail to bring this day together.

Thank you for participating.

Sincerely,

Priya Duggal, PhD, MPH
Director, MD-GEM
Johns Hopkins Bloomberg School of Public Health

David Valle, MD, PHD
Director, MD-GEM
McKusick-Nathans Institute of Genetics Medicine

Dani Fallin, PhD
Associate Director, MD-GEM
Johns Hopkins Bloomberg School of Public Health

Michael Boehnke, Ph.D.

Michael Boehnke is the Richard G. Cornell Distinguished University Professor of Biostatistics and Director of the Center for Statistical Genetics and Genome Science Training Program at the University of Michigan. His research focuses on development and application of statistical designs and analysis methods for human genetics, with emphasis on identification of genetic variants that predispose to human diseases and traits. He is a principal investigator of the FUSION study of the genetics of type 2 diabetes (T2D), steering committee chair of the T2D-GENES multiethnic genome sequencing consortium, and a PI of the Accelerating Medicines Partnership T2D Knowledge Portal project, the BRIDGES bipolar disorder sequencing project, and the InPSYght schizophrenia and bipolar sequencing project. He is a member of the National Academy of Medicine and a fellow of the American Statistical Association and the American Association for the Advancement of Science.

Burroughs Wellcome Fund

The *Burroughs Wellcome Fund* is an independent private foundation dedicated to advancing the biomedical sciences by supporting research and other scientific and educational activities. Within this broad mission, BWF has two primary goals:

- To help scientists early in their careers develop as independent investigators
- To advance fields in the basic biomedical sciences that are undervalued or in need of particular encouragement

BWF's financial support is channeled primarily through competitive peer-reviewed award programs. A Board of Directors comprising distinguished scientists and business leaders governs BWF. BWF was founded in 1955 as the corporate foundation of the pharmaceutical firm Burroughs Wellcome Co. In 1993, a generous gift from the Wellcome Trust in the United Kingdom, enabled BWF to become fully independent from the company, which was acquired by Glaxo in 1995. BWF has no affiliation with any corporation.

http://www.bwfund.org/

Maryland Genetics, Epidemiology and Medicine (MD-GEM) Training Program

The *Maryland Genetics, Epidemiology and Medicine (MD-GEM)* is a pre-doctoral training program that comprehensively integrates Genetics, Epidemiology, and Medicine (GEM). Funded by the Burroughs-Wellcome Fund, the MD-GEM training grant brings together the expertise and training infrastructure of the Johns Hopkins Schools of Public Health and Medicine and the National Human Genome Research Institute. Together, these three institutions can provide laboratory, methodological and clinical expertise and coursework to train the next generation of scientists who can forge new avenues of research and address the rapidly changing field of human genetics. This program trains pre-doctoral students through integration of these important areas by partnering with established mentors and offering integrated learning. We envision a training program that will prepare scientists for the next generation of genetics research.

http://www.hopkinsgenetics.org/

MD-GEM Faculty

Priya Duggal, Co-Director
David Valle, Co-Director
M. Daniele Fallin, Associate Director
Dan Arking
Dimitrios Avramopoulos
Joan E. Bailey-Wilson
Terri Beaty
Aravinda Chakravarti
Debra Mathews
Ingo Ruczinski
Diane M. Becker
Lewis Becker
Larry Brody
Nilanjan Chatterjee
Josef Coresh
Jennifer Deal
Hal Dietz
Andrew Feinberg
Gail Geller
Loyal A. Goff
Ada Hamosh
Kasper Hansen
Julie Hoover-Fong
William Isaacs
Lisa Jacobson
Corrine Keet
Alison Klein
Christine Ladd-Acosta
Jeffrey Leek
Justin Lessler
Brion Maher
Rasika Mathias
Shruti Mehta
Elaine A. Ostrander
Elizabeth A. Platz
Stuart Ray
Debashree Ray
Robert Scharpf
Alan Scott
Margaret Taub
David Thomas
Kala Visvanathan
Jeremy Walston
Xiaobin Wang
Alexander Wilson
Robert Wojciechowski
Peter Zandi

Poster No.	Presenter	Title	Page No.
1	M Atalar	Genetic Architecture of Cystic Fibrosis-Related Diabetes	8
2	Y Xiao	Rare germline variants in known cancer predisposition genes in sporadic chordoma	9
3	K Fletez-Brant	BNBC for Hi-C Data Normalization	10
4	F Chen	Heritability of Familial Pancreatic Cancer using Whole-genome Sequencing Data	11
6	S McClymont	SOX9 ChIP-seq and RNA-seq in PANC-1 cells reveal interesting target genes for pancreatic progenitor biology	12
7	H Ling	A comparison of methods for identification of genetic variants related to age-of-onset of Cystic fibrosis related diabetes	13
8	S Loomis	Exome sequencing analysis of 1,5-AG in the Atherosclerosis Risk in Communities Study	14
9	A Winters	A genome-wide association of genetic determinants of peanut-specific IgG4 at two time points in the Learning Early About Peanut Allergy (LEAP) Study	15
10	J Fu	Detection of de novo copy number deletions from targeted sequencing of trios	16
11	W Kim	Genome-wide association study of emphysema progression	17
12	S Andrews	Placenta DNA methylation is associated with fetal sex at ZNF300	18
13	M Chou	Genotyping Copy Number Variants In Families with Hirschsprung Disease	19
14	JW Fischer	Structured-RNA Decay (SRD)	--
15	K Kanchan	Genomic and structural integrity of human induced pluripotent stem cells	20
16	P Zhang	Exome CNV Overlapping (ECO): an integrative copy number variation caller for exome sequencing	21
17	A Meisner	Evaluating Interactions Between Polygenic Risk Scores and Environmental Risk Factors	22
18	K Stuttgen	Risk Perception Before and After Presymtomatic Genetic Testing for Huntington's Disease: Not What One Might Expect	23
19	DF McKean	Identifying Germline Copy Number Variation in Pancreatic Cancer from a SNP Exome Array	24
20	RJ Longchamps	Exploring the role of Heteroplasmy and Human Disease	25
21	R Mitchell	A Comprehensive Evaluation of the Genetic Architecture of Sudden Cardiac Arrest	26
22	L Boukas	A large subset of the human epigenetic machinery demonstrates co-expression and severe intolerance to loss-of-function variation	27
23	C Peroutka	Retrospective Electronic Medical Record Analysis Identifies a Sizeable Subcohort of Patients at Risk of Hypophosphatasia	28
24	W Li	Maternal Use of Oral Contraceptives Before and After Conception and DNA Methylation Changes in Childhood in The Study to Explore Early Development	29
25	SA Semick	Effects of smoking during pregnancy on the prenatal cortical transcriptome	30
26	H Zhang	Genome-wide association study (GWAS) identifies 19 novel breast cancer loci from analyses accounting for subtype heterogeneity	31
27	B Marosy	Custom Targeted Design Workflow for Next Generation Sequencing	32

Genetic Architecture of Cystic Fibrosis-Related Diabetes

Melis Atalar[1], Briana Vecchio-Pagan[2], Hua Ling[3], Lisa J Strug[4], Rhonda G Pace[5], Harriet Corvol[6], Johanna Rommens[4], Mitchell L Drumm[7], Michael R Knowles[8], Garry R Cutting[1] and Scott M Blackman[1]

[1] Johns Hopkins University School of Medicine, Baltimore, MD
[2] The Johns Hopkins University Applied Physics Laboratory, Laurel, MD
[3] Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD
[4] The Hospital for Sick Children, Toronto, Ontario, Canada
[5] The University of North Carolina at Chapel Hill, Chapel Hill, NC
[6] Pediatric Pulmonology, Hôpital Trousseau, Paris, France
[7] Pediatrics, Case Western Reserve University, Cleveland, OH, United States
[8] Marsico Lung Institute, University of North Carolina, Chapel Hill, NC, United States

Presented by Melis Atalar

Cystic fibrosis (CF) is an autosomal recessive disease caused by mutations in CFTR. CF-related diabetes (CFRD) is an age-dependent complication of CF that affects 19% of adolescents and 40–50% of adults with CF, and is associated with worse lung function, malnutrition and mortality. Wide variation in CFRD onset in people with severe CF was found to be heritable, i.e., attributable to genetic variation outside of CFTR. A genome-wide association study (GWAS) identified CFRD modifier variants intronic and 5′ of SLC26A9, and a candidate association study identified CFRD modifier variants in type 2 diabetes loci (TCF7L2, CDKAL1, CDKN2A/B and IGF2BP2).

To further delineate the genetic architecture of CFRD, the CF modifier consortium genotyped additional patients in a second phase ("GWAS2"). With a combined total of 5,740 CF participants, we conducted a mega-analysis using a Cox proportional hazard regression. Variants at three loci were associated with CFRD onset with genome-wide significance with nearest annotated genes being SLC26A9, TCF7L2 and PTMA. All variants were noncoding, and in each case, associated variants span the annotated gene. Variants near PTMA reached genome-wide significance (eg., rs838455; p:3.8e-8, HR:0.56, 95% CI:0.35-0.75). PTMA encodes a short peptide (thymosine apha 1) recently found to rectify multiple tissue defects in mice with CF and cell lines derived from individuals with CF. PTMA variants are not significant eQTLs for any nearby genes in the pancreas, adipose tissues, brain or muscle (GTEx).

The SLC26A9 variants were reported in the earlier phase (GWAS1) and were supported in the newly recruited individuals (e.g., rs4077468 GWAS1 p-value:3.6e-7, HR:0.67, 95% CI: 0.58-0.77; GWAS2 p-value: 8.9e-3, HR:0.82, %95 CI:0.68-0.99; mega-analysis p-value:2.3e-8, HR:0.72, 95% CI: 0.65-0.81). SLC26A9 encodes for a chloride/bicarbonate exchanger that interacts with CFTR. In addition, in GTEx data these variants are eQTLs for a nearby gene, PM20D1 (rs4077468 eQTL p-values: pancreas: 7.1e-3; adipose: 7.5e-6). PM20D1 encodes a secreted enzyme primarily expressed in the pancreas that regulates N-acyl amino acids, and has been shown to augment energy expenditure.

We reexamined the previously reported T2D risk variants which were associated with CFRD in candidate studies. Variants in TCF7L2 are now genome-wide significant for CFRD (e.g. rs7903146; GWAS1 p: 2.7e-07, HR:1.44, 95% CI: 1.25-1.65; GWAS2 p: 1.9e-6, HR:1.61, 95% CI: 1.32-1.95; mega-analysis p:4.17e-12, HR:1.51, 95% CI:1.34-1.69). TCF7L2 encodes a transcription factor; these intronic variants have been shown in several populations to be associated with type 2 diabetes risk. Variants in CDKAL1, IGF2BP2 and CDKN2B-AS1 were again region-wide significant (p<10e-4). Of note, variants in the CDKN2B-AS1 loci appear to be female-specific (e.g., rs1333045; male-only analysis p: 0.10 HR: 0.91, 95% CI: 0.80-1.02; female-only analysis p:3.2e-7, HR: 0.72, 95% CI:0.64-0.81); combined analysis interaction p-value: 7.0e-3) whereas the T2D association is not sex-specific.

These studies identify multiple genetic modifiers of CFRD onset, which may act through diabetes-specific, CF-specific, or sex-specific pathways. Further delineation of the mechanism of action of these modifiers can inform studies of type 2 diabetes**.**

Content Areas: Human Genetics, Computational Genetics
Keywords: Cystic Fibrosis-Related Diabetes, GWAS, Type 2 Diabetes, Genetic Modifiers

Rare germline variants in known cancer predisposition genes in sporadic chordoma

Yanzi Xiao[1], Bin Zhu, Kristine Jones, Aurelie Vogt, Laurie Burdette, Wen Luo, Belynda Hicks, Neal Freedman, Stephen Chanock, Dilys Parry, Alisa Goldstein and Xiaohong Yang

[1]Human Genetics Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health

Presented by Yanzi Xiao

Chordoma is a rare bone cancer that occurs in the skull base and spine with an incidence rate less than 0.1 per 100,000 in the United States. We previously identified germline T gene duplications as a major susceptibility mechanism in familial chordoma. Common and rare germline variants in the T gene were also found to be associated with both familial and sporadic chordoma. However, underlying genetic susceptibilities for the majority of chordoma cases remain unknown. To systematically characterize rare germline variants in established cancer predisposition genes in chordoma, we extracted whole exome sequencing (WES) data from 133 unrelated sporadic chordoma cases recruited from the United States and Canada for 114 known cancer predisposition genes. Rare variants were defined as <0.1% in the 1000 Genomes Project, Exome Sequencing Project, and ≤ 2 families from our in-house database of 1,000 cancer-prone control families. Among the 114 known cancer predisposition genes, we identified 172 rare nonsynonymous variants of which 13 are loss-of-function (LOF) variants. To assess the overall genetic burden of rare germline variants, we performed a rare variant burden test (SKAT statistics) comparing the chordoma cases to 598 unrelated population controls from two cohort studies (Cancer Prevention Study [CPS]-II and Prostate, Lung, Colorectal and Ovarian Screening Trial [PLCO]) that were sequenced using the same platform as the chordoma cases. Cases had increased frequencies of rare germline variants for MET, UROD, ERCC4, TRIM37, and BRCA2 compared to controls (p-value < 0.01). After Bonferroni correction, the MET gene remained significant (p-value < 4x10-4). Of particular interest, a missense variant in MET (chr7:116397716) was shared by 4 chordoma cases but none of the controls. This variant, which was not reported in any public or in-house control datasets, is predicted to be deleterious by all 13 in silico programs we evaluated. Our results suggest that rare germline variants in MET may be associated with susceptibility to sporadic chordoma.

Content Area: Human Genetics
Keywords: Chordoma, Exome sequencing, rare variants

BNBC for Hi-C Data Normalization

Kipper Fletez-Brant, MHS, MS and Kasper Daniel Hansen, PhD

Presented by Kipper Fletez-Brant

Current methods for Hi-C data normalization operate on single samples and correct for biases between cells in one contact map. We desire to compare a given cell from a contact map across multiple samples and show that doing so requires removing unwanted variation. Using Hi-C data from multiple samples we characterize this unwanted variation and demonstrate that it varies between cells in the contact map. We develop the method BNBC to correct this unwanted variation and show that it meaningfully improves cross-sample comparisons.

Content Area: Other
Keywords: statistics, hi-c, normalization, genomics

Heritability of Familial Pancreatic Cancer using Whole-genome Sequencing Data

Fei Chen[1], Nicholas Roberts[2] and Alison P. Klein[1-3]

[1] Department of Epidemiology, the Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
[2] Department of Pathology, the Sol Goldman Pancreatic Cancer Research Center, the Johns Hopkins Medical Institution, Baltimore, MD 21231, USA
[3] Department of Oncology, the Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institution, Baltimore, MD 21231, USA

Presented by Fei Chen

Pancreatic cancer is the 3rd leading cause of cancer deaths in the US, and is projected to become the 2nd by 2020. We have demonstrated that inherited genetic factors are in important component of risk. More than 19 independent loci have been identified though GWAS studies, and several high-penetrance genes have been identified including BRCA2, BRCA1, PALB2, ATM, CDKN2A as well as mis-match repair genes. Having a family history of pancreatic cancer is one of the strongest risk factors for the disease, and 5-10% of newly diagnosed patients report a close relative with pancreatic cancer. The heritability of pancreatic cancer has not been well estimated. A population-based twin study in Europe has estimated the heritability for pancreatic cancer to be 36% (95% CI = 0-53%). However, such estimate may be biased upwardly due to shared environment. In our study, we applied the restricted maximum likelihood (GREML) approach to estimate heritability from a case-control cohort using whole-genome sequencing (WGS) data. Our study population is a combined cohort of 658 familial pancreatic cancer (FPC) patients (defined as kindreds with 2 first-degree relatives with pancreatic cancer) and 809 participants of Alzheimer's Disease Neuroimaging Initiative (ADNI). The case and control cohort were whole-genome sequenced at Illumina. After quality control, genotype data on 1,322 individuals for 14,055,912 autosomal single-nucleotide variants (SNVs) were included in our heritability analysis. Preliminary heritability estimates will be presented.

Content Area: Genetic Epidemiology
Keywords: heritability, pancreatic cancer, whole-genome sequencing

SOX9 ChIP-seq and RNA-seq in PANC-1 cells reveal interesting target genes for pancreatic progenitor biology

Sarah A McClymont[1], Hannah E Edelman[1], Andrew S McCallion[1] and Michael J Parsons[1,2]

[1] McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD
[2] Department of Developmental and Cell Biology, University of California, Irvine, CA

Presented by Sarah A McClymont

Both Type I and Type II diabetes eventually result in a paucity of functioning beta cells. Since there are associated cost and donor scarcity problems with islet transplantation, we are focused on studying the endogenous creation of new beta cells. We found that specialized intercalated-duct cells called centroacinar cells (CACs) contribute to the regeneration of beta cells in zebrafish. CACs exist in humans, but their role in regeneration is controversial. In zebrafish, we found that the transcription factor (TF) Sox9b is necessary for maintaining CAC progenitor status – haploinsufficient fish regenerate beta cells faster after ablation than their wildtype siblings. Thus, to elucidate how zebrafish CACs are able to regenerate beta cells, we are interested in the transcriptional targets of Sox9b.

Content Area: Human Genetics
Keywords: Diabetes, Regeneration, SOX9, ChIP-seq, RNA-seq

A comparison of methods for identification of genetic variants related to age-of-onset of Cystic fibrosis related diabetes

Hua Ling[1], Peng Zhang[1], Elizabeth W Pugh[1], Melis Atalar[2] and Scott Blackman[2]

[1] Center for Inherited Disease Research, Johns Hopkins Genomics
[2] The McKusick-Nathan Institute of Genetic Medicine

Presented by Hua Ling

Cystic fibrosis (CF) is a monogenic disease that affects more than 80,000 people worldwide and causes life-limiting lung disease and pancreatic dysfunction. Diabetes (CF-related diabetes or CFRD) is the most common extrapulmonary complication of CF and affects >40% people with CF by adulthood with a broad range of age of onset. This variation in CFRD risk has been shown to be heritable, and Blackman et al. (2013) identified five loci associated with CFRD onset, analyzed as a survival trait while excluding related individuals in a total of 3,059 samples. To better account for relatedness in survival analyses, we investigate different strategies using a family subset of the above data, the CF Twin and Sibling study, which includes 396 samples from 288 small families (siblings and half siblings) genotyped on the Illumina 610-Quad. Our preliminary analyses show using mixed-effect Cox models, either with family-specific random intercept or with correlated random intercept using a kinship coefficient matrix, yield slightly better control of inflation of type 1 error compared to Cox proportional hazard model including related individuals ($\lambda = 1.00$ and 1.096 for family-specific and correlated random intercept respectively vs 1.12 for Cox proportional hazard model with related individuals), but not compared to maximally unrelated subset analysis ($\lambda = 1.03$). Analyses of PC-adjusted Martingale residuals in linear mixed model with relatedness as random effects yield similar results ($\lambda = 1.096$). Supported by CF Foundation.

Content Area: Genetic Epidemiology
Keywords: Cystic fibrosis, Genome wide association analysis, Survival analysis, Mixed effect model, Relatedness

Exome sequencing analysis of 1,5-AG in the Atherosclerosis Risk in Communities Study

Stephanie Loomis[1], Priya Duggal[1], Elizabeth Selvin[1,2], Josef Coresh[1,2], Adrienne Tin[1], Eric Boerwinkle[3], James Pankow[4] and Anna Kottgen[1,5]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
[2] Welch Center for Prevention, Epidemiology, & Clinical Research, The Johns Hopkins University, Baltimore MD
[3] Department of Epidemiology, The University of Texas Health Science Center at Houston School of Public Health at Houston, Houston, TX
[4] Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN
[5] Institute of Genetic Epidemiology, Medical Center and Faculty of Medicine - University of Freiburg, Freiburg, Germany

Presented by Stephanie Loomis

Introduction: 1,5-anhydroglucitol (1,5-AG) has been gaining interest as a nontraditional biomarker of type 2 diabetes. Recent genome-wide association studies (GWAS) have identified several associated common variants, but impact of rare coding variants has yet to be evaluated using a targeted assay of 1,5-AG in European and African ancestries.

Methods: We performed exome sequencing analysis on 1,5-AG among European ancestry (N=6,589) and African ancestry (N=2,309) participants without diagnosed diabetes in the Atherosclerosis Risk in Communities (ARIC) study. Both single variant and gene-based (T1 and SKAT) tests were done.

Results: 15 variants reached genome-wide significance with 1,5-AG (range 3.4-32.8 ug/mL) among European ancestry individuals, and four of these replicated in the African ancestry sample. Three of these variants are on chromosome 17 in or near SLC5A10 (top SNP: rs148178887, p=1.13E-36, beta= -10.38) and one variant is on chromosome 2 in or near LCT and UBXN4 (rs1050115, p=5.69E-09, beta= -0.80). SLC5A10 was significantly associated with 1,5-AG in both the European ancestry (beta=-8.65, p=2.5E10-114) and African ancestry (beta=-2.99, p=7.5E10-07) samples (Bonferroni corrected significance threshold=2.0E10-06) using SKAT. SLC5A10 was also associated with 1,5-AG using the T1 gene-based analyses (p=2.8E10-55), and replicated in the African ancestry sample (p=6.5E10-03).

Conclusions: We identified several rare, coding variants in individuals of European ancestry and replicated them in an African ancestry sample, confirming the importance of regions identified by previous studies near SLC5A10 and LCT for 1,5-AG. The large effect sizes of these variants indicate a strong impact on of variants in these regions on 1,5-AG levels.

Content Area: Genetic Epidemiology
Keywords: "1,5-AG", type 2 diabetes, hyperglycemia biomarkers, exome sequencing

A genome-wide association of genetic determinants of peanut-specific IgG4 at two time points in the Learning Early About Peanut Allergy (LEAP) Study

Alexandra Winters[1], Meher Boorgula[2], Claire Malley[1] and Rasika Mathias[1]

[1] Johns Hopkins School of Medicine, Department of Allergy & Clinical Immunology
[2] University of Colorado, Department of Medicine

Presented by Alexandra Winters

The LEAP study showed a protective effect of early peanut exposure for infants at high risk of developing peanut allergy. Peanut-specific IgG4 increased over time in peanut consumption as compared to avoidance subjects, but the degree to which it improved differed between subjects in the peanut consumption group. In this study, we undertake whole genome sequencing (WGS) to understand the genetic determinants of clinical outcomes in the LEAP Study. First, we leverage genome-wide genotype array (GWA) data generated as part of our WGS and perform a GWAS on 267 participants in the peanut consumption group in the LEAP study. Tests for association were performed on all single nucleotide variants (SNVs) that passed quality control and age, sex, and the first five principal components for ancestry were included in the model.

The peak association signal for peanut-specific IgG4 at 60 months of age (the conclusion of the LEAP trial) was located on chromosome 2. The peak SNPs are common intronic variants with minor allele frequencies >10% and are expression quantitative trait loci (eQTLs) for the SEPT2 gene in a number of tissues. An additional association signal was seen for peanut-specific IgG4 at 60 months of age on chromosome 6 mapping to a region between HLA-DQA1 and HLA-DQB1, and is an eQTL for HLA-DQB1 in multiple tissues.

Our results are promising, implicating two candidate genes, SEPT2 and HLA-DQB1. SEPT2 has previously been shown to regulate airway epithelial barrier function, and deficiencies in the epithelial barrier are a known cause of allergy. Additionally, previous work has shown association with variants in HLA-DQB1 and increased risk of peanut allergy independent of asthma. We are currently analyzing all sequence-identified variants to follow up on this early result and to identify addition rare and novel variants.

Content Area: Human Genetics
Keywords: peanut allergy, IgG4, HLA, SEPT2

Detection of de novo copy number deletions from targeted sequencing of trios

Jack M Fu[1], Elizabeth J Leslie[2], Alan F Scott[3], Jeffrey C Murray[4], Mary L Marazita[5], Terri H Beaty[6], Robert B Scharpf[7] and Ingo Ruczinski[1]

[1] Department of Biostatistics, Johns Hopkins University, Baltimore, MD
[2] Department of Human Genetics, Emory University, Atlanta, GA
[3] Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD
[4] Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA
[5] School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA
[6] Department of Epidemiology, Johns Hopkins University, Baltimore, MD
[7] Department of Oncology, Johns Hopkins University, Baltimore, MD

Presented by Jack Fu

De novo copy number deletions have been shown to be associated with many diseases. However, no method currently exists that explicitly leverages targeted sequencing data from case-parent trios to identify de novo copy number deletions. Here we present our algorithm Minimum Distance for Targeted Sequencing (MDTS). MDTS exhibits comparable sensitivity (recall), but a much lower false positive rate compared to currently available CNV callers when applied to targeted sequencing trios. MDTS is available as open source software developed for R at www.github.com/jmf47/MDTS.

Content Area: Statistical Genetics
Keywords: de novo copy number variants; case-parent trios; targeted sequencing; oral clefts

Genome-wide association study of emphysema progression

Woori Kim[1], Margaret Parker[2], Phuwanat Sakornsakolpat[2], Michael Cho[2,3], Edwin Silverman[2,3] and Terri Beaty[1]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[2] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA
[3] Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA

Presented by Woori Kim

Introduction: Chronic obstructive pulmonary disease (COPD) and one of its main sub-types, emphysema, represent a complex and heterogeneous pulmonary disease that includes emphysema, a progressive form of COPD reflecting of normal lung parenchyma. Genome-wide association studies have identified genetic factors associated with cross-sectional measures of emphysema, but genetic determinants of longitudinal change in quantitative measures of emphysema remain largely unknown. Our study aims to identify the genetic influences of emphysema progression measured by computed tomography (CT) imaging in current and ex-smokers from the COPDGene study, a longitudinal cohort enrolling Non-Hispanic Whites (NHW) and African Americans (AA) between ages of 45 and 80 years at baseline with a minimum of 10 pack-year smoking history

Methods: A total of 4,238 subjects (3,078 NHW subjects and 1,160 AA subjects) with complete genotype and CT data in the COPDGene were included in the analysis. Two key quantitative measures of emphysema based on CT imaging (percent emphysema and adjusted lung density) were tested for genetic associations in subjects followed over 5 years. Emphysema progression was defined as the difference in CT emphysema measures between phase 1 and 2. Imputed genome-wide markers generated using the Haplotype Reference Consortium panel as a reference panel were used to identify markers associated with these phenotypes. Multiple linear regression was performed under an additive genetic model in a genome-wide association study (GWAS). The analysis was stratified by race (i.e. separately for NHW and AA). Baseline measures of age, pack-years of smoking, smoking status and emphysema (percent emphysema and adjusted lung density accordingly) and sex were considered as covariates. Principal components were adjusted for potential population stratification within each racial group.

Results: We identified two loci associated with progression of percent emphysema at genome-wide significance among AA subjects. Markers in or near LOC100133091 on chromosome 7 (rs182836571; minor allele frequency (MAF) =3.6%; P=7.74e-09) and ST6GALNAC3 on chromosome 1 (rs115187551; MAF = 1.3%; P=2.13e-08) showed significant evidence of influencing change in percent emphysema.

Conclusions: Our study provides evidence for a role of genetic variants with low frequency in contributing to emphysema progression among subjects with African ancestry. Our study identified new genomic loci associated with emphysema progression. These findings may point that development and progression of emphysema are distinctly influenced by different genetic determinants. Further validation of these associations in larger African populations is required to elucidate the biologic pathways of emphysema and COPD.

Content Area: Genetic Epidemiology
Keywords: genetics, emphysema, emphysema progression, chronic obstructive pulmonary disease

Placenta DNA methylation is associated with fetal sex at ZNF300

Shan V. Andrews[1,2], Christine Ladd-Acosta[1-3], Kelly M. Bakulski[4], Jason I. Feinberg[2,5], Rakel Tryggvadottir[3], Lisa A. Croen[6], Irva Hertz-Picciotto[7,8], Craig J. Newschaffer[9,10], Ruofan Yao[11], Carolyn M. Salafia[12], Andrew P. Feinberg[3,13], Kasper D. Hansen[3,14,15] and M. Daniele Fallin[2,3,5]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
[2] Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins Bloomberg School of Public Health
[3] Center for Epigenetics, Institute for Basic Biomedical Sciences, Johns Hopkins School of Medicine
[4] Department of Epidemiology, University of Michigan School of Public Health
[5] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health
[6] Division of Research, Kaiser Permanente Northern California
[7] Department of Public Health Sciences, School of Medicine, University of California Davis
[8] MIND Institute, University of California Davis
[9] AJ Drexel Autism Institute, Drexel University
[10] Department of Epidemiology and Biostatistics, Drexel University Dornsife School of Public Health
[11] Department of Obstetrics, Gynecology and Reproductive Medicine, University of Maryland School of Medicine
[12] Placental Analytics LLC
[13] Department of Medicine, Johns Hopkins School of Medicine
[14] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
[15] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine

Presented by Shan Andrews

Background: In normal development, there are sex-specific differences in placenta function and responses to nutritional, stress-related, or environmental insults. The placenta also undergoes vast epigenomic reprogramming. While a few gene expression studies have interrogated the mechanism of these sex differences, only one genome-wide study of placenta DNA methylation (DNAm) differences by fetal sex has been performed. This study was limited in its coverage of the placenta methylome and to address this limitation, we used a comprehensive measurement of methylation (whole-genome bisulfite sequencing) on 37 (Nmales = 17, Nfemales = 20) fetal side placenta samples and searched for differentially methylated regions (DMR) according to fetal sex.

Results: We identified 1 genome wide significant (ppermutation = 0.015) DMR where males exhibited on average 15% higher methylation levels in a CpG island promoter of ZNF300, a gene previously associated with cell proliferation and tumorogenesis. This fetal sex DMR is placenta-specific, and is consistently replicated with similar magnitude in 7 additional sets of samples (made up of 6 independent studies), including full term and preterm placenta datasets and mixed cell and single cell type datasets, and persists in different disease states. Females can be dichotomized into those who exhibit ZNF300 methylation levels similar to males and those that have lower methylation levels and drive the DMR. Methylation levels in this DMR are associated with placenta perimeter (p = 0.044), area (p = 0.009) and maximum diameters (p = 0.059); this finding is also driven the lower methylation female group. Finally, there is suggestive evidence that ZNF300 DMR methylation levels mediate chromosome 5 and chromosome X SNPs influences on these morphological features.

Conclusions: Placenta methylation levels at the ZNF300 promoter are differential by sex and associated with placenta morphological features. Both of these findings are driven by a subset of female samples that have lower ZNF300 methylation levels than males and the rest of female samples.

Content Area: Computational Genetics
Keywords: placenta, DNA methylation, fetal sex, whole genome bisulfite sequencing

Genotyping Copy Number Variants in Families with Hirschsprung Disease

Michael Chou[1], Robert B. Scharpf[2], Priya Duggal[1], Terri Beaty[1] and Aravinda Chakravarti[3]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
[2] Department of Biostatistics , Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
[3] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA

Presented by Michael Chou

Hirschsprung disease (HSCR) is a highly heritable complex birth defect, and rare deleterious coding variants in 24 different genes have been identified as influencing risk. Previous studies have also demonstrated different karyotypic anomalies in HSCR patients, and we have recently identified an enrichment of large deletions and duplications in at least 15 different chromosomal regions. Smaller copy number variants (CNVs) may also increase HSCR risk but these remain poorly studied. To address this gap in HSCR, we analyzed 295 HSCR cases and their families (n = 450), including 74 trios using Omni 2.5 single nucleotide polymorphism (SNP) arrays. To identify CNVs, we extend a Bayesian mixture model for copy number inference in case-control studies to trio-based study designs. Key to our approach is a Mendelian model for transmission of copy number alleles from parents to offspring. Markov chain Monte Carlo is used to derive probabilistic estimates of copy number in HSCR trios.

Content Area: Genetic Epidemiology
Keywords: Hirschsprung Disease, Copy Number Variant

Genomic and structural integrity of human induced pluripotent stem cells

Kanika Kanchan[1], Claire Malley[1], Lisa R. Yanek[1], Linzhao Cheng[1], Zack Wang[1], Diane Becker[1], Lew Becker[1], Ingo Ruczinski[2] and Rasika Mathias[1]

[1] Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD
[2] Department of Biostatistics, Bloomberg School of Public health, Johns Hopkins University, Baltimore, MD

Presented by Kanika Kanchan

Human induced pluripotent stem cells (hiPSCs) constitute model systems with enormous capability to shed light on the pathogenesis and mechanisms for a wide range of human diseases. Their application as disease models depends on the genomic integrity and stability through their generation. Recent studies confirm various genomic instabilities, such as, chromosomal aberrations, copy number variations (CNVs) and differences in single nucleotide polymorphisms (SNPs) in hiPSCs. We analysed genome-wide genotype array data (Illumina Infinium MEGA Chip) on 1,321,228 genetic variants to evaluate genomic instability of 132 hiPSC lines generated from their matched peripheral blood mononuclear cells using non-integrating episomal vectors. We observed very low levels of genotype mismatch between genetic variants available on the genotype array between the 132 donor and hiPSC pairs (mean mismatch error rate = 0.004%). Processed SNP array data were subjected to automated CNV calling followed by manual inspection of all CNVs. For automated analysis, we applied Hidden Markov Models on the genotype data to identify CNVs in the samples using R package 'VanillaICE'. Given prior documentation of high rates of false positive CNV calls through automated pipelines, sub-chromosomal plots were generated from the log R ratios and B-allele frequencies for manual inspection and confirmation of the automated calls. Our automated pipeline detected no differences in average numbers of CNVs between the hiPSC and donor DNA. There was an average of 7 deletions and 3 amplifications per hiPSC line, and 6 deletions and 3 amplifications per donor sample. All called CNVs were then compared within the hiPSC and donor pair to identify non-overlapping CNVs (i.e. those called in the hiPSC but not the donor, or those that overlap but were of differing lengths between the hiPSC and donor). Manual inspection of non-overlapping CNVs revealed that most pairs (N=130) had no detectable difference in CNVs called in the hiPSC from those in the donor DNA. Two hiPSC lines had two different CNVs as compared to the donor: one line had a deletion, and other had amplification. Our results suggest that much of the called CNVs in the hiPSC are pre-existing in the donor DNA. Further, there are only low levels of genomic instability in our lines as quantified by CNV differences between the hiPSC and its paired donor DNA.

Content Area: Human Genetics
Keywords: hiPSCs, CNVs, GWAS, LRR, BAF

Exome CNV Overlapping (ECO): an integrative copy number variation caller for exome sequencing

Peng Zhang[1], Hua Ling[1], Elizabeth Pugh[1] and Kim Doheny[1]

[1] Center for Inherited Disease Research (CIDR), Johns Hopkins Genomics, Institute of Genetic Medicine, The Johns Hopkins School of Medicine

Presented by Peng Zhang

Due to the uneven distribution of reads and the sparse nature of target regions for whole exome sequencing (WES), calling copy number variations (CNVs) has been a challenge and most of existing programs can only use read counts as inputs and calls often vary between programs. For example, the numbers of CNVs called were 174,183 (ExomeDepth), 2,670 (HMZDelFinder), and 38,952 (XHMM), respectively, for ~1600 WES samples from the Centers for Mendelian Genomics (CMG) project. As part of the validation process, we found that some confirmed causal CNVs were called by multiple programs while others were not. In addition, each program often requires different input files and its output format often varies with different breakpoints for the CNV calls, which makes it difficult to compare and summarize results across programs.

We present here a practical pipeline that integrates multiple CNV calling programs and generates one combined VCF-like report with merged calls and annotations. It incorporated three prevalent CNV calling programs (ExomDepth [Plagnol et al. 2012], CANOES [Backenroth et al. 2014], and CODEX [Jiang et al. 2015]) with the ability to incorporate results from two additional programs (XHMM [Fromer and Purcell 2014] and HMZDelFinder [Gambin et al. 2017]). In addition, our pipeline: 1) Generates read counts only once, either from BAM or CRAM; 2) Runs the three methods in parallel; 3) Merges calls by a user-defined overlap percentage and a size threshold; 4) Provides annotation such as gene names in the regions and call frequencies.

Content Area: Computational Genetics
Keywords: Copy number variations, Whole exome sequencing, Statistical genetics, Computational genetics

Evaluating Interactions Between Polygenic Risk Scores and Environmental Risk Factors

Allison Meisner[1] and Nilanjan Chatterjee[1]

[1] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Presented by Allison Meisner

The identification of interactions between genetic and environmental risk factors has been pursued in a variety of areas over the past several decades. In 1997, the case-only design for studying gene-environment interactions was proposed by Umbach and Weinberg. This approach has been utilized broadly in the intervening decades and remains popular today. At the same time, researchers have recognized that most traits are the result of a large number of genetic variants acting together, motivating consideration of polygenic risk scores (PRS). PRSs capture variation across a number of genetic risk factors, in many cases offering improvements in prediction over individual variants. There is now growing interest in identifying interactions between PRSs and environmental risk factors. To that end, we have proposed a case-only approach to evaluating interactions between a PRS and an environmental risk factor. Provided certain assumptions, such as independence between the PRS and the environmental risk factor, are met, the relative risk interaction parameter can be estimated using standard linear regression methods. We use simulations to demonstrate gains in efficiency of over 50% for our method compared to a traditional case-control analysis and illustrate application of our method using breast cancer data from the UK Biobank.

Content Area: Statistical Genetics, Genetic Epidemiology
Keywords: Interaction, polygenic risk score, case-only

Risk Perception Before and After Presymtomatic Genetic Testing for Huntington's Disease: Not What One Might Expect

Kelsey Stuttgen[1,2], Rachel Dvoskin1, Juli Bollinger[1], Allison McCague[1,2], Barnett Shpritz[3], Jason Brandt[3,4] and Debra Mathews[1]

[1] Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD
[2] Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
[3] Department of Psychiatry and Behavioral Science, Johns Hopkins University School of Medicine, Baltimore, MD
[4] Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD

Presented by Kelsey Stuttgen

In 1983, Huntington's Disease (HD) was the first genetic disease mapped using DNA polymorphisms. Shortly thereafter, presymptomatic genetic testing for HD began in the context of two clinical trials. One of these trials, the John's Hopkins HD presymptomatic testing protocol (JH HD presymtomatic protocol), began in 1986 and enrolled 180 individuals during the first 10 years of the program.

As part of the presymptomatic protocol, participants were asked to indicate on a line from 0% to 100% his or her perceived risk of developing HD. These risk perception values were collected at 16 time points before and after testing and results disclosure. The last time point was collected as long as 11 years after disclosure of genetic test results. The current study investigated changes in risk perception scores before and after genetic testing was performed on individuals enrolled in the JH HD presymtomatic protocol.

The data reveal a wide range of perceived risk, even in the presence of a conclusive test result. While the risk perception scores of most individuals change in the way one would expect, meaning risk perception decreased after receiving a negative test result or increased after receiving a positive test result, several participants demonstrated unexpected changes in risk perception after disclosure. These unexpected changes include increased risk perception after receiving a negative test result, decreased risk perception after receiving a positive test result, and no change in risk perception after receiving a test result.

The data suggest that individuals' perception of their risk of disease is influenced by more than merely the results of genetic testing. This finding is important for genetic counselors and healthcare providers, as it suggests that even fairly comprehensive patient education and disclosure of genetic test results may not ensure that an individual fully appreciates their risk of disease. Further study is required to understand what are the other factors influencing individuals' risk perception.

Content Area: Human Genetics
Keywords: Genetic Testing, Bioethics, Presymtomatic Testing, Risk Perception

Identifying Germline Copy Number Variation in Pancreatic Cancer from a SNP Exome Array

David F. McKean[1], Stephen Cristiano[1,2], Jacob Carey[3], Ingo Ruczinski[2], Alison P. Klein[1,3,4] and Robert Scharpf[2] for the Pancreatic Cancer Case-Control Consortium

[1] Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland, USA
[2] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland
[3] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
[4] Department of Pathology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

Presented by David F. McKean

Pancreatic cancer is the currently 3rd leading cause of cancer death in the United States. Inherited genetic factors play an important role in pancreatic cancer risk. While common single nucleotide variants have been examined in large-scale genome-wide association studies of familial pancreatic cancer, copy number variants have been less extensively studied. With the goal of characterizing the contribution of germline copy number alterations to pancreatic cancer risk using data from 3,974 cases and 3,624 controls in the Pancreatic Cancer Case Control Consortium genotyped in the IlluminaOmniExpress Exome Array at the Center for Inherited Disease Research, we developed a pipeline for identifying both common and rare copy number variants from this platform. We used hidden markov models to identify copy number variants in the individual samples from normalized probe intensities. For known copy number polymorphic regions and regions commonly altered in our study, we re-genotyped the collection of 7,598 samples using (1) a conventional Bayesian mixture model fit to all samples and (2) a novel multi-batch mixture model where chemistry plate and other known batch metadata were modeled hierarchically. Bayes factors were used to compare non-nested models and posterior predictive distributions were used to assess the adequacy of these models. Through these analyses, we identified regions with common copy number alterations. Importantly, we also identify regions that would have been erroneously genotyped as polymorphic if the batch metadata were ignored. The mixture models are implemented in the R package CNPBayes available from Bioconductor.

Content Area: Statistical Genetics
Keywords: Pancreatic Cancer, Exome Array, Copy Number Variants

Exploring the role of Heteroplasmy and Human Disease

Ryan Longchamps[1,2], Yun Soo Hong[3],Megan Grove[4], Eric Boerwinkle[4], Eliseo Guallar[3] and Dan Arking[1]

[1] Mckusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD
[2] Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
[3] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[4] Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX

Presented by Ryan Longchamps

Mitochondria play a critical role in energy metabolism and have been linked to human disease and mortality. Several biological process have been hypothesized to explain the role of mitochondrial dysfunction in disease, such as declines in energy production, altered rates of apoptosis, and elevated free radical production. Previous reports have shown mitochondrial DNA (mtDNA) may be the key component underlying mitochondrial dysfunction, and thus human disease. Our lab has recently shown mtDNA quantity, also known as mtDNA copy number (mtDNA-CN) is associated with coronary artery disease (CAD), cardiovascular disease (CVD), sudden cardiac death (SCD), and overall mortality. However, little is known about how an accumulation of mtDNA mutations, resulting in increased heteroplasmy, affect the quality of mtDNA. We hypothesized as the heteroplasmic content of an individual's mtDNA increased; mitochondrial dysfunction would increase resulting in human disease.

Utilizing whole genome sequence data from 3,600 participants of the Atherosclerosis Risk in Communities (ARIC) cohort we tested our hypotheses. Heteroplasmy were called using mitoAnalyzer – a software package specifically developed to analyze mtDNA sequence data. Heteroplasmy was called as any site with at least 500X mitochondrial DNA coverage and 3% of reads with the alternative allele. Individuals were removed if at least 10,000 sites did not contain 500X coverage. We developed a mtDNA quality score whereby we summed the total number of heteroplasmic sites for an individual and divided by the total number of sites interrogated. The final mtDNA quality score represented the standardized residuals from a linear model adjusting for age, sex, and mtDNA-CN.

Surprisingly, Initial analyses revealed the number of heteroplasmic sites to be protective for incident CHD, incident CVD, and mortality (P = 7.81x10-4, 6.37x10-4, and 0.015 respectively). These results remain consistent across several heteroplasmic alternative allele fraction cutoffs including 1%, 5% and 10%. These results contradict other published data, and we hope further analyses in the coming month will further elucidate the relationship between heteroplasmy and disease.

Content Area: Human Genetics, Genetic Epidemiology
Keywords: Heteroplasmy, Cardiovascular Disease, Coronary Heart Disease, Mortality

A Comprehensive Evaluation of the Genetic Architecture of Sudden Cardiac Arrest

Rebecca Mitchell[1], Foram Ashar[1,2], Dan E. Arking[1] and Nona Sotoodehnia[3]

[1] Institute of Genetic Medicine, Johns Hopkins, Baltimore, USA
[2] CHARGE SCD Working Group
[3] Cardiovascular Health Research Unit, Division of Cardiology, Departments of Medicine and Epidemiology, University of Washington

Presented by Rebecca Mitchell

Background. Sudden cardiac arrest (SCA) accounts for 10% of adult mortality in Western populations. While several risk factors are observationally associated with SCA, the genetic architecture of SCA in the general population remains unknown. Furthermore, understanding which risk factors are causal may help target prevention strategies.

Methods. We carried out a large genome-wide association study (GWAS) for SCA (n=3,939 cases, 25,989 non-cases) to examine common variation genome-wide and in candidate arrhythmia genes. We also exploited Mendelian randomization methods using cross-trait multi-variant genetic risk score associations (GRSA) to assess causal relationships of 18 risk factors with SCA.

Results. No variants were associated with SCA at genome-wide significance, nor were common variants in candidate arrhythmia genes associated with SCA at nominal significance. Using cross-trait GRSA, we established genetic correlation between SCA and (1) coronary artery disease (CAD) and traditional CAD risk factors (blood pressure, lipids, and diabetes), (2) height and BMI, and (3) electrical instability traits (QT and atrial fibrillation), suggesting etiologic roles for these traits in SCA risk.

Conclusions. Our findings show that a comprehensive approach to the genetic architecture of SCA can shed light on the determinants of a complex life-threatening condition with multiple influencing factors in the general population. The results of this genetic analysis, both positive and negative findings, have implications for evaluating the genetic architecture of patients with a family history of SCA, and for efforts to prevent SCA in high-risk populations and the general community.

Content Area: Human Genetics
Keywords: Sudden cardiac arrest, genome-wide association study, mendelian randomization

A large subset of the human epigenetic machinery demonstrates co-expression and severe intolerance to loss-of-function variation

Leandros Boukas, James Havrilla, Aaron Quinlan, Hans Bjornsson and Kasper Hansen

Presented by Leandros Boukas

We define a set of 295 human genes whose protein products contain domains classifying them as writers/erasers/readers of DNA methylation, histone methylation or acetylation, or as chromatin remodelers, which we collectively call the epigenetic machinery. Systematic exploration of these genes reveals the versatility of the readers, which can have enzymatic and/or multiple reading functions. Despite the fact that many encode for enzymes, a large subset of these genes are highly intolerant to loss-of-function variation, even when compared to transcription factors. This intolerance is primarily driven by the protein domains that directly mediate the epigenetic function, as revealed by within-gene comparisons to other domains. Interestingly, more than a third of those genes demonstrate co-expression within many tissues, with the degree of co-expression being strongly related to variation intolerance. Finally, the co-expressed subset is highly enriched for genes associated with neurological dysfunction, even when accounting for dosage sensitivity. These findings reveal key characteristics and highlight the essentiality of the human epigenetic machinery, while concomitantly providing candidates for future disease gene discovery.

Content Area: Computational Genetics
Keywords: Epigenetics

Retrospective Electronic Medical Record Analysis Identifies a Sizeable Subcohort of Patients at Risk of Hypophosphatasia

Christina Peroutka, MD[1]; Adekemi Alade, MD, MPH[1]; Mark Marzinke, PhD[2]; John McGready, PhD[3]; Natasha Parikh, MS[1]; Kerry Schulze, PhD[3]; and Julie Hoover-Fong, MD, PhD, FACMG[1]

[1] The Johns Hopkins University, McKusick-Nathans Institute of Genetic Medicine
[2] The Johns Hopkins University, Department of Pathology Clinical Chemistry
[3] The Johns Hopkins University, School of Public Health

Presented by Christina Peroutka

Introduction: Hypophosphatasia (HPP) is an ultra-rare, metabolic condition with pleiotropic anifestations including perinatal lethality or later-onset fractures, early tooth-loss, osteoporosis, seizures, chronic bone/muscle pain, fatigue and/or respiratory compromise. Caused by variants in tissue-nonspecific alkaline-phosphatase(TNSALP/ALPL), HPP is characterized by low serum alkaline-phosphatase(AP). AP is commonly ordered, but low values are often unrecognized as pathologic by healthcare providers. Given a broad disease spectrum and under-recognized indicator, HPP is likely underdiagnosed. Misdiagnosis as idiopathic osteoporosis leads to treatment with bisphosphonates, which counteract available HPP-specific asfotase-alpha enzyme replacement therapy. We conducted a risk-prevalence study for HPP by query of a large, hospital-based electronic medical record(EMR), demonstrating that EMR-query may be used to identify individuals at risk of rare disease.

Methods: The Johns Hopkins(JH) IRB approved this protocol under a waiver of consent. JH patients from all clinical sites with >1 AP level reported from Jan,2013-Dec,2015 were compiled by medical record number(MRN). Site and ICD-9 codes were merged to MRNs with >1 low AP based on age- and sex-specific norms. MRNs with liver disease (direct bilirubin >5), renal disease(creatinine >2.6), or hematologic, oncologic, drug and/or transplant codes associated with low AP were excluded. Patients with 5-20 AP levels, at least 80% abnormal, were prioritized for comprehensive review. EMR data extracted included medical history, family history, and physical features suggestive of HPP. A REDCap database was maintained on a password-protected, institutionally-supported server. Results were generated using STATA(V13,College Station,TX).

Results: There were 983,458 AP levels from 156,459 patients, with 45,233(22,175 patients <19yo) low AP levels from 11,730 patients. Exclusion by ICD-9 and site-codes yielded 7,129 patients for review. Assuming patients with 5-20 AP levels over 3 years, >80% abnormal, are at increased risk of HPP, 351 patients were eligible for chart review (157 with 100% abnormal AP levels). Data from 200 patients is presented. Medical history is suggestive of HPP in 62(31%) patients: 6 of 200(3%) with premature tooth-loss, 18(9%) excessive caries, 16(8%) nephrolithiasis, 19(9.5%) muscle pain, and 67(33.5%) chronic pain. 16(8%) have osteoporosis; 6(3%) received bisphosphonates and 1 received a PTH-analog. Multiple specialists provide care: 28(14%) endocrinology, 39(19.5%) orthopedics, 18(9%) rheumatology, 19(9.5%) genetics, 10(5%) nephrology, and 9(4.5%) pain-management. FH was recorded as follows: in notes 164/200(82%), in EMR FH table 75(38%), for >3 relatives 95(48%), as a pedigree 17(9%). FH was consistent with HPP in 4(2%). HPP was not on the differential for patients with low AP, though 24(12%) had molecular testing for other conditions.

Conclusions: This study demonstrates that EMR-query of a large patient population can identify a subcohort with a high likelihood of having HPP that warrant further investigation. Despite patients in this subcohort reaching out to various medical specialists, and reporting symptoms consistent with HPP, HPP is rarely, if ever, considered as a diagnosis. We are contacting patients through a separate prospective study to offer genetic counseling, molecular testing, and treatment to those affected. This type of study serves as a proof of principle that EMR-related prevalence studies may be applied broadly to improve identification and treatment of patients with rare conditions, including HPP.

Content Area: Human Genetics
Keywords: Bioinformatics, Clinical History, Databases, Delineation of Diseases, Metabolic Disorder

Maternal Use of Oral Contraceptives Before and After Conception and DNA Methylation Changes in Childhood in The Study to Explore Early Development

Weiyan Li[1], B.K.Lee[2], N. Gidaya[2], C. Newschaffer[2], L. A. Schieve[3], D. E. Schendel[4], N.Jones[5,6], G. C. Windham[7], L. A. Croen[8], A.P. Feinberg[9], C. Ladd-Acosta[1] and M.D. Fallin[10]

[1] Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD
[2] Department of Epidemiology and Biostatistics and the A.J. Autism Institute, Drexel University School of Public Health, Philadelphia, PA
[3] National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA
[4] Department of Public Health, Institute of Epidemiology and Social Medicine, Aarhus University, Aarhus, Denmark, Department of Economics and Business, National Centre for Register-based Research, Aarhus University, Aarhus, Denmark, and Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Denmark
[5] Biomedical Research Informatics Core, Michigan State University, East Lansing, MI
[6] Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC
[7] California Dept of Public Health, Richmond, CA
[8] Kaiser Permanente Division of Research, Oakland, CA
[9] Center for Epigenetics, Johns Hopkins School of Medicine, Baltimore, MD
[10] Department of Mental Health, Johns Hopkins School of Public Health, Baltimore, MD

Presented by Weiyan Li

Autism spectrum disorder (ASD) is a devastating neurodevelopmental syndrome that affect up to 1 in 68 children. The highly skewed male to female ratio (4 to 1) consistently observed in ASD gives rise to the extreme male brain theory, which proposes that characteristics of ASD are presentations of male-typical characteristics on the extreme end of a spectrum. Estrogen plays an important role in masculinization of the brain and sex-dimorphic behavior, and has been linked to ASD in both animal studies and human population. In the US, over 10 million women are exposed to highly potent synthetic estrogen and progesterone through use of oral contraceptives (OC). Failures in contraception lead to unintended exposure of highly potent synthetic estrogen to over a half million fetuses. DNA methylation is an important regulatory mechanism of gene regulation and is responsive to environmental stimuli, and can serve as either a biomarker for previous environmental exposures that are difficult to assess, and a mediation mechanism between environmental stimuli and consequential disease outcomes. To the best of our knowledge, no study has examined the impact of maternal use of oral contraceptives in relation to DNA methylation profile in children. In this study, we will examine the epigenome of 902 children aged 2-5 years old to search for DNA methylation signatures related to maternal use of oral contraceptives.

Content Area: Genetic Epidemiology
Keywords: Autism spectrum disorders, estrogen, oral contraceptives, DNA methylation

Effects of smoking during pregnancy on the prenatal cortical transcriptome

Stephen A. Semick[1], Leonardo Collado-Torres[1,2], Christina A. Markunas[3], Joo Heon Shin[1], Amy Deep-Soboslay[1], Ran Tao[1], Laura J. Bierut[4], Brion S. Maher[5], Eric O. Johnson[6], Thomas M. Hyde[1,7,8], Daniel R. Weinberger[1,7-10], Dana B. Hancock[3], Joel E. Kleinman[1,7] and Andrew E. Jaffe[1,2,5,10,11]

[1] Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, 21205, USA
[2] Center for Computational Biology, Johns Hopkins University, Baltimore, MD, 21205, USA
[3] Behavioral and Urban Health Program, Behavioral Health and Criminal Justice Division, RTI International, Research Triangle Park, NC, 27709, USA
[4] Department of Psychiatry, Washington University School of Medicine, St Louis, MO 63110, USA
[5] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA
[6] Fellow Program and Behavioral Health and Criminal Justice Division, RTI International, Research Triangle Park, NC, 27709, USA
[7] Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA
[8] Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, 21205, USA
[9] Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, 21205, USA
[10] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA
[11] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA

Presented by Stephen A. Semick

Cigarette smoking during pregnancy is a major public health concern. While there are well described consequences in early child development, there is very little known about the effects of maternal smoking on human cortical biology during prenatal life. We therefore performed a genome-wide differential gene expression analysis using RNA sequencing (RNA-seq) on prenatal (N=33; 16 nicotine-exposed) as well as adult (N=207; 57 active smokers) human post-mortem prefrontal cortices. Nicotine exposure during the prenatal period was directly associated with differential expression of 14 genes: in contrast, during adulthood, despite a much larger sample size, only 2 genes showed significant differential expression (FDR<10%). Moreover, 1,315 genes showed significantly different exposure effects in the pre- versus post-natal human cortex largely driven by prenatal differences that were enriched for pathways previously implicated in nicotine addiction and synaptic function. These genes were further enriched for genes implicated in syndromic autism spectrum disorder (ASD) and significantly differentially expressed as a set in brains of postmortem patients with ASD compared to controls. Lastly, to better untangle the effects of nicotine exposure and nicotine-predisposing genetic variants in the brain, we integrated expression quantitative trait loci to uncover novel cis-regulatory effects for nicotine dependence on local transcription in the adult human brain (N=237), including associations with COMMD7 (p=1.56×10-9), CYP2T2P (p=1.17×10-7), and AXL (p=4.76×10-5). These results underscore the enhanced sensitivity to the biological effect of nicotine in the developing brain and offer novel insight into the effects of nicotine exposure's on the prenatal human brain and the independent functional consequences of genetic risk variants for nicotine dependence. They also begin to address the relationship between in utero exposure to nicotine and the heightened risk for the subsequent development of neuropsychiatric disorders.

Content Area: Human Genetics, Genetic Epidemiology
Keywords: maternal smoking during pregnancy, brain development, autism spectrum disorder, transcriptomics

Genome-wide association study (GWAS) identifies 19 novel breast cancer loci from analyses accounting for subtype heterogeneity

Haoyu Zhang[1], Thomas Ahearn[2], Ni Zhao[1], Nilanjan Chatterjee[1,3] and Montserrat Garcia-Closas[2]

[1] Johns Hopkins University, Bloomberg School of Public Health, Department of Biostatistics
[2] National Cancer Institute, Division of Cancer Epidemiology & Genetics
[3] Johns Hopkins University, School of Medicine, Department of Oncology

Presented by Haoyu Zhang

Background: Breast cancer is a highly heterogeneous disease with different subtypes having different etiology, clinical behaviors, and genetic risk factors. Currently, most common breast cancer risk loci identified in GWAS were discovered using standard case-control regression or ER-negative/triple negative vs control regression. These approaches are less powered to identify SNPs when tumor heterogeneity is present. We propose a novel mixed-effect two-stage logistic regression model and conduct a GWAS to identify risk loci that simultaneously accounts for multiple correlated tumor characteristics.

Method: We analyzed data collected from 81 studies of the Breast Cancer Association Consortium with a total of 108,946 cases and 96,201controls of European descent to identify loci associated with breast cancer risk. Genotype data were measured using two genotyping platforms (ICOGs and OncoArray) and imputed using reference panel from 1000 Genome project. Separate analyses were conducted based on ICOGs and OncoArray samples and combined in a meta-analysis. SNPs with a minor allele frequency < 0.01 and those within 500Kb of or correlated at coefficient of determination ()>0.1 with known risk SNPs were excluded from analysis. We applied a novel mixed effect two-stage logistic regression model to assess the overall and subtype-specific association between a SNP and breast cancer, with cancer subtype being defined by tumor grade and biomarkers ER, PR, HER2.
This method efficiently accounts for multiple testing, correlation between markers and missing tumor marker data.

Results: We identified 19 novel loci to be genome-wide significantly (P<5x10-8) associated with overall breast cancer risk. Of them, heterogeneity tests showed 5 loci had heterogeneity only modified by ER, 5 loci had heterogeneity only modified by Grade, 1 loci had heterogeneity only modified by HER2. The other 8 loci were associated with overall disease or were heterogeneous for more than one marker. Conditional analyses further identified another 12 SNPs to be significantly associated with breast cancer risk in global test (P<1x10-5) after adjusting for the top signals in the 19 regions.

Conclusion: Using methods that account for tumor heterogeneity, we identified 19 novel breast cancer risk loci.

Content Area: Human Genetics
Keywords: Subtypes Heterogeneity, Two-stage model, Tumor characteristics

Custom Targeted Design Workflow for Next Generation Sequencing

Beth Marosy[1], David Mohr[1], Kimberly Doheny[1] and Alan Scott[1]

[1] Johns Hopkins Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University

Presented by Beth Marosy

Genome Wide Association (GWA) and linkage studies that identify a chromosomal region(s) may require additional follow-up sequencing in order to find a causal variant(s) that contributes to the genetic trait. Targeted sequencing is a powerful tool for querying regions of interest in large sample cohorts at high depth for relatively low cost compared to whole exome or whole genome sequencing. However inefficient design of custom probes across regions of homology and repetitive elements can cause a decrease in the selection metrics which lead to increased sequencing costs, rendering custom capture as an expensive option. By utilizing a rigorous design workflow, these challenges can be addressed prior to manufacturing of a custom product, aiding in reducing unforeseen sequencing costs. Here we have developed a workflow for designing custom targeted panels and present cases where improved sequencing conditions were met. Using the UCSC genome browser, genes and genomic elements of interest are identified for inclusion in the design. GALAXY provides tools to extract exons, add flanks, merge or subtract regions and calculate base coverage. Probes are designed across regions of interest in a tiered fashion, reducing stringency parameters for each subsequent pass of uncovered target regions not covered by a probe in a previous pass. Sequences of the probes from each pass are analyzed using BLAT to identify any potentially problematic probes that will affect selection due to homology. Threshold settings within BLAT can be varied according to the project needs, scaling the stringency of the analysis to either increase coverage or balance sequencing costs. Application of BLAT analysis can improve sequencing selection metrics by >20%. Depending on the capture size, this could prevent as much as 50% of added sequencing costs due to inefficient design.

Content Area: Human Genetics
Keywords: Targeted Sequencing, Next Generation Sequencing

Epigenetic alterations in childhood reflect prenatal exposure to maternal infection

Martha Brucato[1], Shan V. Andrews[1], Yinge Qian[2], Gayle Windham[3], Diana Schendel[4], Laura Schieve[5], Craig Newschaffer[6], Andrew Feinberg[1], Lisa Croen[2], M. Daniele Fallin[1] and Christine Ladd-Acosta[1]

[1] Johns Hopkins University, Baltimore, Maryland, USA
[2] Kaiser Permanente Northern California Division of Research, Oakland CA, USA
[3] California Department of Public Health, Richmond, CA, USA
[4] Aarhus University and National Centre for Register-based Research, Aarhus, Denmark
[5] Centers for Disease Control and Prevention, Atlanta, GA, USA
[6] Drexel University, Philadelphia, PA, USA

Presented by Martha Brucato

Prenatal exposure to maternal immune activation (MIA), particularly infections and fever, has been linked with altered neurodevelopment in the exposed offspring, yet we have a limited understanding of causal mechanisms. To explore the potential biological consequences of prenatal exposure to maternal infections, we examined 929 children, aged 2-5 years, in the Study to Explore Early Development, phase I (SEED I) with both genome-scale whole blood DNA methylation data, from the Illumina 450K array, and in utero infection exposure data, ascertained via structured maternal phone interview. We used linear models, adjusted for cell type composition, sex, ancestry, and other unwanted variation via surrogate variables, to identify differentially methylated loci associated with prenatal exposure to infection. We found one site in an intergenic region on chromosome 5 that was significantly (q-value = 0.005) hypomethylated in children whose mothers had an infection during the preconception period. We also identified 2 genomic loci, within the IQSEC1 and EPS8L3 genes, showing significant decreases in DNA methylation (q-value=0.014 for IQSEC1 and q-value = 0.036 for EPS8L3) among children whose mothers had an infection during the third trimester. The differences in percent methylation increased in magnitude when comparing children whose mothers reported infections in every trimester of pregnancy (n=56) to those whose mothers reported no infections during pregnancy (n=589). This may reflect a dose-response relationship between a cumulative prenatal infection exposure and methylation at the identified sites. Although we detected these differences in blood, reference datasets indicate that methylation at the intergenic locus on chromosome 5 is strongly correlated across blood and brain. This site is also predicted to be near an enhancer-like region in human astrocytes. IQSEC1 is thought to be involved in synaptic transmission, as both a scaffolding and signaling protein. It is highly expressed in brain tissues, particularly the frontal cortex, as well as whole blood. EPS8L3 is not well studied but is likely involved in actin regulation, which is important in neuronal structures like the postsynaptic density and dendritic spine. Our findings suggest that epigenetic changes related to prenatal infection exposure can present in childhood samples, and also provide candidate loci for studies examining potential epigenetic mediation of prenatal MIA exposure and atypical neurodevelopment.

Content Area: Genetic Epidemiology
Keywords: epigenetics, prenatal infection, autism spectrum disorder

Upstream regulatory element(s) increase expression of SLC26A9 leading to a delayed age at onset of diabetes in cystic fibrosis

Anh-Thu Lam[1], Melis Atalar[1], Briana Vecchio-Pagan[1], Scott Blackman[1] and Garry Cutting[1]

[1] Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD

Presented by Anh-Thu Lam

Diabetes is an age-dependent complication of Cystic Fibrosis (CF) that is highly heritable and affects 40-50% of the adult CF population. CF-related diabetes (CFRD) is associated with worse lung function, malnutrition and mortality, and is clinically and histologically distinct from type 1 and type 2 diabetes. Genetic association studies have identified significant association with CFRD for variants near and within the SLC26A9 gene. SLC26A9 encodes a chloride/bicarbonate transporter that interacts with CFTR, the protein defective in CF, and is an interesting candidate as it has also been identified as a modifier of intestinal obstruction and drug response in CF. A biologic role for SLC26A9 in exocrine pancreatic function is supported by single-cell RNA sequencing indicating that SLC26A9 and CFTR are co-expressed in pancreatic ductal cells. To assess the genetic architecture of SLC26A9, the entire locus was sequenced in 762 F508del homozygote individuals with CF. Two linkage disequilibrium (LD) blocks were identified and defined by a single recombination in intron 8. Two common haplotypes (24.2% and 28.4% frequency) exist within the 5' LD block containing variants associated with higher-risk (HR) and lower-risk (LR) of diabetes, respectively. No coding variants were in LD with either haplotype, and no individual variant showed greater association with diabetes than the two common haplotypes. To determine if the region containing the diabetes-associated variants encompasses regulatory sequences, the 2.3 kb upstream of SLC26A9, corresponding to the HR and LR haplotypes, was cloned into a dual luciferase/renilla reporter (DLR) system. Combined analysis of the normalized data from 3 independent transfections into PANC-1, a surrogate for pancreatic ductal cells, with 4 biological clones per haplotype (technical replicates: N=71 for LR and N=72 for HR) showed that HR had a 13% lower promoter activity compared to LR (p=9.31E-9). Luciferase expression was negligible when the DNA sequences were cloned in the reverse orientation indicating that the 2.3kb region 5' of SLC26A9 drives expression in an orientation-specific manner. These results are consistent with eQTL predictions from RNA sequencing data suggesting that LR variants in the 2.3 kb region correlate with higher levels of expression than HR variants (rs4077469; p=2.25E-08; odds ratio=0.72). To further investigate this region, we bisected the 2.3 kb into 1.173 kb, which is further from exon 1 and contains three key CFRD-associated SNPs, and 1.172kb, which is closest to exon 1 and contains two key CFRD-associated SNPs. Promoter activity was exclusively observed in the 1.172 kb fragment (~9-fold increase compared to baseline controls), whereas, no activity was observed for the 1.173 kb region as detected by the DLR assay in PANC-1. Interestingly, no differences in activity were observed between the 1.172 kb fragment bearing LR and HR variants, suggesting that the 1.173 kb fragment may contain regulatory element(s) that function to influence the significant difference in expression observed with the 2.3 kb. Taken together, these results indicate that increased expression of SLC26A9, likely influenced by upstream regulatory element(s), may delay the onset of diabetes in individuals with CF, possibly by modification of pancreatic ductal function.

Content Area: Human Genetics, Molecular Genetics
Keywords: Cystic Fibrosis-Related Diabetes, Cis Regulatory Elements, Luciferase, GWAS, Genetic modifiers

Diverging Genome-Wide Neuronal DNA Methylation at Base-Resolution Across Human Brain Development

Amanda J. Price[1,2], Leonardo Collado-Torres[1], Nikolay A. Ivanov[1], Joo Heon Shin[1], Ran Tao[1], Emily Burke[1], Wei Xia[1], Liang Ma[1], Yankai Jia[1], Thomas M. Hyde[1,3,4], Joel E. Kleinman[1,4], Daniel R. Weinberger[1-5] and Andrew E. Jaffe[1,6,7]

[1] Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA
[2] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[3] Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA
[4] Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD, USA
[5] Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, USA
[6] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[7] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Presented by Amanda Price

DNA methylation (DNAm) plays an integral role in cell identity and brain development. Previous studies have identified widespread DNAm changes across human brain development, but all have used homogenate tissue. To characterize the base-resolution DNAm landscape over human cortical development within cell types, we isolated neuronal and glial-enriched populations from 24 human dorsolateral prefrontal cortex (DLPFC) samples aged 0-23 years using fluorescence-activated nuclear sorting (FANS) and performed whole-genome bisulfite sequencing (WGBS).

We confirmed that neurons and glia differed significantly in their DNAm landscapes (11,179 cell-type differentially methylated regions (DMRs), FWER<5%; 4.82M differentially methylated CpGs, FDR<5%). These DMRs were replicated in independent WGBS data (98.4% concordant, ρ=0.925; Lister et al., Science 2013), enriched for expected brain-relevant gene sets/pathways, and functionally validated using complementary FANS-derived chromatin accessibility and RNA sequencing.

We found more age-associated DMRs within rather than across cell type (2,179 versus 129 DMRs, FWER<5%), with largely increasing neuronal DNAm and decreasing glial DNAm over development. These DMRs were strongly enriched for brain transcribed and enhancer sequences compared to non-brain tissues (>8 fold). The rate of change per decade of life was 1.5 times higher in neurons than glia (p<2.2e-16). Importantly, ~40% of these changes would not have been observed in homogenate tissue. Using unsupervised clustering, infant neuronal DNAm was more similar to glia than older neurons at these DMRs, which were significantly associated with biological processes involved with neuronal arborization and synapse maturation.

We confirmed greater non-CpG DNAm levels in neurons than glia (7.68M sites, FDR<5%) that increased over postnatal development (3.19M sites, FDR<5%). We observed that neuronal and glial DNAm levels were very similar in infant life, and neuronal cells diverged in their DNAm levels across aging. This phenomenon was seen in both CpG and non-CpG contexts at both sets of developmentally regulated genomic loci, further suggesting a general accumulation of DNAm. By leveraging cell type-specific and developmental RNA sequencing data, we also elucidate the role that non-CpG methylation plays in regulating cortical gene expression.

This work highlights the dynamic DNAm landscape of developing cell types in the postnatal human cortex, including the accumulation of CpG and non-CpG DNAm that distinguishes young from more developed neurons, and further refines the temporal dynamics of acquiring (non-)CpG DNAm in neurons and the interplay between CpG and non-CpG DNAm within genomic loci across brain development. The divergence of DNAm from early infanthood highlights the important regulatory cascades guiding human brain development using epigenetic mechanisms specific to neuronal cells that may be dysregulated in neuropsychiatric disease.

Content Areas: Human Genetics, Molecular Genetics, Other
Keywords: brain development, DNA methylation, WGBS, epigenomics, neuropsychiatric disease

Integrative Analysis of Two RNA-seq Dataset to Improve Understanding of Biological Mechanism of Brain Development

Cristian Valencia[1], Leonardo Collado-Torres[2], Emily Burke[2] and Andrew Jaffe[2]

[1] Johns Hopkins Bloomberg School of Public Health
[2] Lieber Institute for Brain Development

Presented by Cristian Valencia

Introduction: New evidence suggests that brain disorders have a neurodevelopmental component. Brain development is a very complex process that requires a tightly coordinated gene expression dynamic at both temporal and spatial dimensions. Deficiencies at any stage of this process can increase the risk of disease. Understanding the transcriptional changes across different ages in non-psychiatric individuals may help to elucidate the biological mechanism of brain disorders. Also, studying changes in gene expression in non-psychiatric samples may prevent confounding with factors associated with schizophrenia and bipolar disorder like medication, drug use, cigarette smoking, etc.

The Lieber Institute for Brain Development (LIBD) has mRNA sequence of 363 non-psychiatric individuals and 2 brain regions: Dorsolateral prefrontal cortex (DLPC) and hippocampus. Also, the BrainSpan Atlas of the Developing Human Brain project has mRNA sequence of 40 control individuals and 16 brain regions including DLPC and hippocampus. Both projects follow different technologies and pipelines which makes comparison difficult. Interestingly, eighteen individuals were sequenced in both projects. Thus, we aim to evaluate the feasibility of integrating both datasets by re-analyzing the data and comparing concordance between these overlapping samples.

Methods: Raw data from the LIBD and BrainSpan project will be re-aligned to the same reference human genome GRch38 following the same pipeline. Briefly, quality of the read was analyzed using FASTQC. Sequencing adaptors were removed using Trimmomatic software. Reads were aligned using HISAT, an aligner program designed to improve popular TopHat2 program. Then, Samtools and RSEQC were used to calculate quality metrics for alignment. Finally, expression level will be evaluated using featureCounts software. Several metrics from aligned reads like quality metrics, gene coverage, junction annotation, and gene and exon level expression will be calculated for every sample.

Dataset integration will be evaluated using the concordance of the eleven overlapping individuals and the DLPC region. Concordance will be assess using summary metrics of the FASTQ files at both base and read level and alignment-level summary statistics. The relative ratio of same-individual-different-dataset variability to different-individual-same-dataset variability will be calculated. Also, the concordance between gene RPKMs for same-individual-different-dataset to different-individual-same-dataset will be calculated to confirm that reprocessing of data has improved the possibility to combine the two datasets.

Results: The BrainSpan project consisted of 40 unique donors and 607 brain samples. Overall, mapping rate ranged from 35% to 87% with a mean of 70%. The mitochondrial mapping ranged from 0.6% to 68% with a mean of 16%. Three samples were removed due to potential swapping and 30 samples were also removed due to high mitochondrial mapping rate. The concordance among the two datasets will be further discussed.

Conclusions: Integration of both datasets will help to fully characterized transcriptional activity across human brain development.

Content Area: Genetic Epidemiology
Keywords: RNAseq, neurodevelopment, transcriptomic

Family Based Association Tests of Myopia reveal a potentially hidden association signal upstream of two GABA receptor genes

Candace Denise Middlebrooks[1], Claire L. Simpson[2], Anthony M. Musolf[1], Laura Portas[3], Federico Murgia[3], Elise Ciner[4], Dwight Stambolian[5] and Joan E. Bailey-Wilson[1]

[1] National Human Genome Research Institute, National Institutes of Health, Baltimore, MD, United States
[2] Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN, United States
[3] Institute of Population Genetics, CNR, Li Punti, Sassari, Italy
[4] Salus University, Elkins Park, PA, United States
[5] Ophthalmology-Stellar Chance Lab, University of Pennsylvania, Philadelphia, PA, United States

Presented by Candace Middlebrooks

Myopia is an eye condition in which the light entering the eye does not focus on the retina resulting in distant objects appearing out of focus. Within recent years, the incidence and prevalence of myopia have increased in most populations and has reached epidemic proportions in several Asian countries. We have performed a family based association study using Exome Chip genotyping (Illumina Human Exome v1.1 array plus 24,263 custom SNPs) in five family cohorts for a total of 1718 subjects in 261 families. These cohorts include Amish, Ashkenazi Jewish, African American, Caucasian and Chinese American families who have multiple individuals affected with myopia.

Individuals in the families were defined as myopic if their average refractive error was <= -1 Diopter (D) and were considered unaffected if their average refractive error was > 0.0 D. Children were considered unaffected as follows: MSE>=+2D (ages 6-10) or MSE>=+1.5D (ages 11-20). After quality control, there were ~127,000 polymorphic SNPs available for analysis. Both gene-level and single-variant association analyses were performed using Family Based Association Test (FBAT) software. This resulted in a significant signal in a novel region upstream of two gamma-Aminobutryric Acid (GABA) receptor genes (GABRA6; GABRB2). GABA is a neurotransmitter that has previously been implicated in refractive development. The associated SNP, rs1373602, is not found in the Genotype-Tissue Expression (GTEx) project, but a nearby SNP, rs62381591, has been identified as an expression quantitative trait locus for the GABRA6 gene. As the significant variant is common (~48% across populations), we wondered why the larger, population-based associations studies of Myopia have not found a signal in this regions. Upon further analysis, we learned that this variant is not in high Linkage disequilibrium (LD) with any other variants in our dataset (highest r2 was ~0.002) and is indicated as triallelic in the 1000 genomes dataset (although it was biallelic in our smaller dataset).

Hence, this triallelic SNP may be filtered out before GWAS and there may not be another SNP that tags this region. We plan to follow-up this analysis by collaborating with groups that have performed genome-wide genotyping studies of Myopia to determine if this signal was missed due to the aforementioned reasons.

Content Area: Statistical Genetics
Keywords: Genetics, Myopia, Statistics, FBAT, Exome